



VII EDYCJA E-LEARNING FUSION  
DIGITAL LEARNING & training market  
CONFERENCE & EXHIBITION 2023

NEW

14-15 Listopada 2023  
**WARSZAWA**

**Kacper Łodzikowski**  
Vice President of AI Capabilities, Pearson

Zaprezentuję temat:

**Implementing AI in training & development processes**

Organizator:

Digital  
Learning  
Centre

# AI Wave 1: Expertise-driven

**Write rules for AI, and it will follow them** (e.g. GUIDON tutoring system, Clancey 1984)

**\*\* I THINK THAT THE ORGANISMS PRESENT ARE DIPLOCOCCUS-PNEUMONIAE, PSEUDOMONAS-AERUGINOSA, AND NEISSERIA-MENINGITIDIS.**

Good. The age of the patient gives evidence for concluding that Diplococcus-Pneumoniae might be present (Rule 507). The fact that the patient has been seriously burned is good reason to suspect Pseudomonas-Aeruginosa (Rule 578). Both conclusions apply because the infection which requires therapy is meningitis, no organisms were seen on a gram stain of the culture, and the type of the infection is bacterial.

The age of the patient also leads one to suspect Neisseria-Meningitidis (Rule 507). Do the fact that the infection was acquired while the patient was hospitalized and information about the diseases to which the patient has been exposed support this conclusion?

**\*\* NO**

Right, this would be good reason to rule out Neisseria-Meningitidis (Rule 561).

What facts about this case tell you that the organisms that might be present are: E.Coli and Klebsiella-Pneumoniae?

**\*\* HELP**

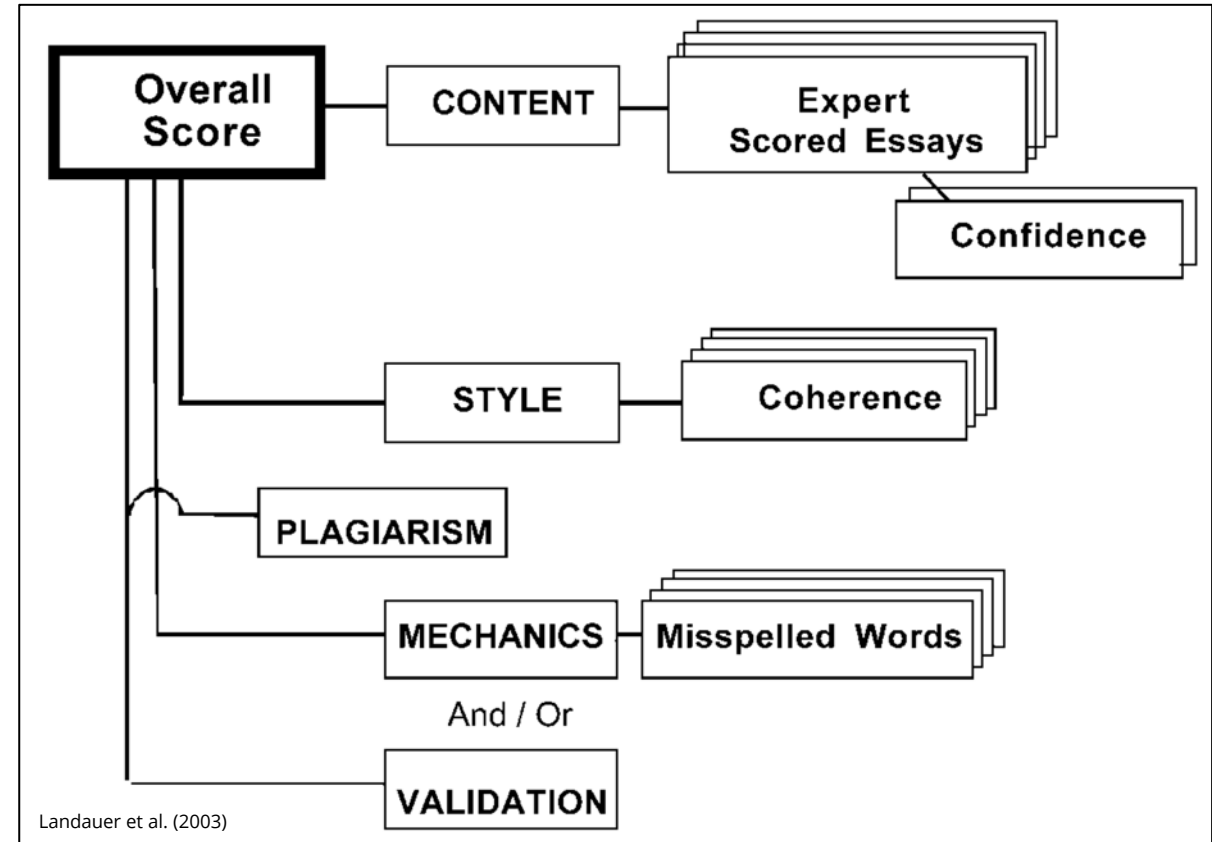
**RULE578**

IF: 1) The infection which requires therapy is meningitis, and  
2) Organisms were not seen on the stain of the culture, and  
3) The type of the infection is bacterial, and  
4) The patient has been seriously burned

THEN: There is suggestive evidence (.5) that pseudomonas-aeruginosa is one of the organisms (other than those seen on cultures or smears) which might be causing the infection

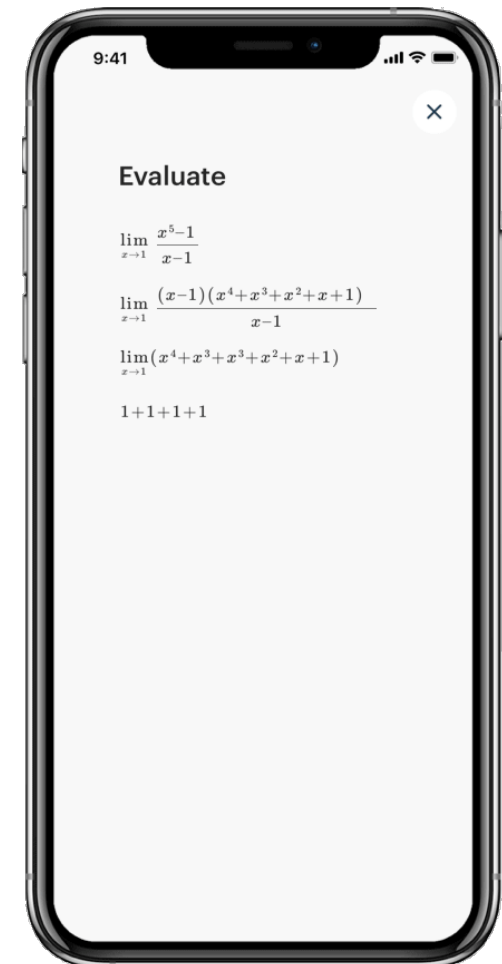
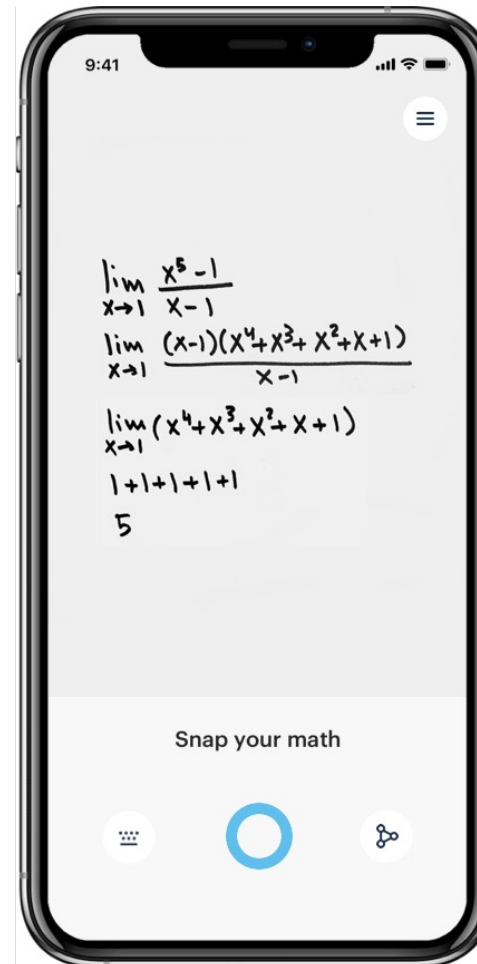
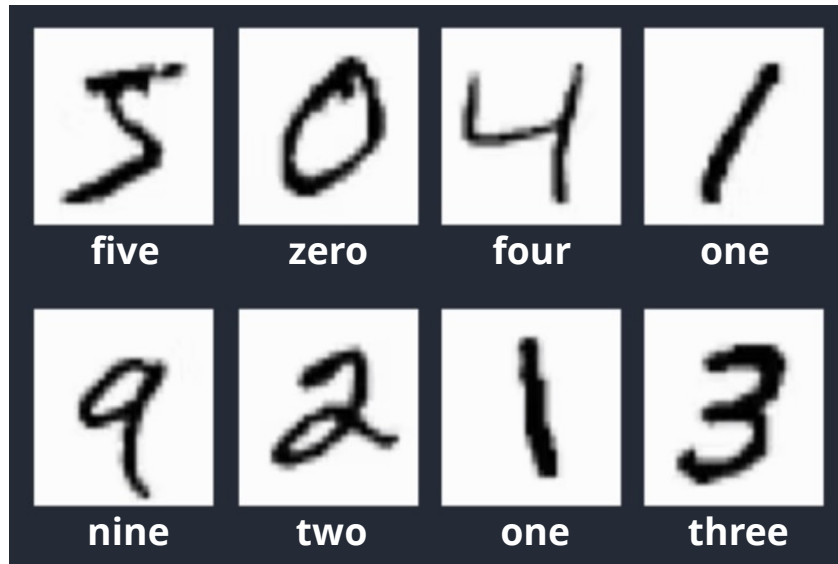
# AI Wave 2: Data-driven (machine learning)

Collect data & use expertise to teach AI to mimic humans (e.g. Intelligent Essay Assessor, Foltz et al. 1999)



# AI Wave 3: Data-driven (deep learning)

**Collect labelled data & neural network will learn w/o expertise** (e.g. Aida Calculus, Pearson 2019)



# AI Wave 3.5: Data-driven (generative deep learning)

Collect unlabelled data & neural network will learn w/o expertise (e.g. Karpathy 2015)

Neural network learns to generate text in the style of Tolstoy's *War and Peace* by reading it...

100 times

tyntd-iafhatawiaoahrdemot lytdws e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e  
plia tklrqd t o idoe ns,smtt h ne etie h,hregtrs nigtike,aoaenns lng

500 times

we counter. He stutn co des. His stanted out one ofler that concossions and was  
to gearang reay Jotrets and with fre colt ofp paitt thin wall. Which das stimn

700 times

Aftair fall unsuch that the hall for Prince Velzonski's that me of  
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort

2,000 times

"Why do what that day," replied Natasha, and wishing to himself the fact the  
princess, Princess Mary was easier, fed in had oftended him.



# Primary value of AI in education is in personalisation

## Just-in-time tutoring

The screenshot shows a chemistry problem-solving interface. On the left, a problem is presented:  $4\text{Ag(s)} + \text{O}_2\text{(g)} \rightarrow 2\text{Ag}_2\text{O(s)}$ . The text describes a teapot that has become completely tarnished and asks for the number of moles of silver in a 751g teapot, to be expressed to three significant figures. A hint button is visible. Below the problem, a student has entered the answer  $\Delta m = -98 \cdot 10^{-11} \text{ g}$ . The interface indicates the answer is incorrect, with 5 attempts remaining. A chat window on the right contains the text: "To help you work through this problem, I will ask you a series of questions, guiding you step by step. One mole of silver is equivalent to 107.9 g of silver. The solid silver teapot weighs 751g. How many moles of silver is this? Express your answer to three significant figures and include the appropriate units." Below the chat is a text input field with the placeholder "Or ask a question..." and a "Send" button.

## Practice activity generation

The screenshot shows a genetics lesson titled "Dominantly Inherited Disorder". The text explains that although many harmful alleles are recessive, some disorders are dominant, such as achondroplasia. It mentions that heterozygous individuals have the dwarf phenotype (Figure 14.18) and that 99.99% of the population do not have achondroplasia. A Punnett square diagram is shown for a cross between a dwarf phenotype ( $Dd$ ) and a normal phenotype ( $dd$ ). The diagram shows the parents' genotypes, the resulting sperm and egg cells, and the possible offspring genotypes ( $Dd$  and  $dd$ ). A caption below the diagram reads: "Figure 14.18 Achondroplasia: a dominant trait." An "AI Study Tools" overlay is visible on the right side of the screen. It includes a "Beta" badge, a close button, and a button to "Can you generate some practice questions?". Below this, there are options for where to pull questions from: "Current section" (selected), "Whole chapter", and "What kind of questions are you looking for?": "Multiple choice" (selected), "Short answer". There is a "Send" button and a text input field with the placeholder "Type a message". At the bottom of the overlay are "New topic" and "Send" buttons.



At the heart of personalisation lies assessment

## Three foundations of assessment efficacy:

### **Validity**

Measure what is required for making claims

### **Fairness**

Not biased against any learner group

### **Reliability**

Consistent results across testing occasions

# Assessment validity

Measure what is required for making claims

What you can measure easily / initially



What you eventually want to measure





# Assessment validity

Measure what is required for making claims

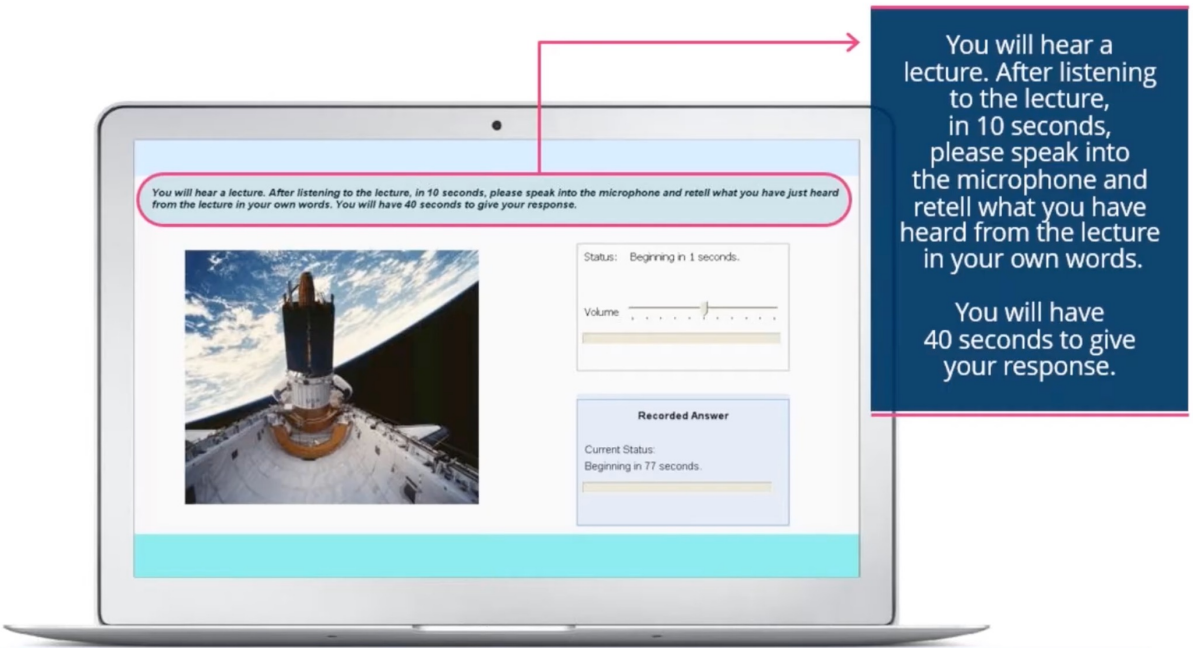
## What you can measure easily / initially

Read the text. Select a word or phrase to fill the gap.

Turn off your mobile phone now.  
The use of mobile phones and other electronic devices is not \_\_\_\_ during the examination.

A  accepted  
B  recommended  
C  permitted

## What you eventually want to measure



You will hear a lecture. After listening to the lecture, in 10 seconds, please speak into the microphone and retell what you have just heard from the lecture in your own words. You will have 40 seconds to give your response.

Status: Beginning in 1 seconds.

Volume: [Slider]

Recorded Answer

Current Status: Beginning in 77 seconds.

You will hear a lecture. After listening to the lecture, in 10 seconds, please speak into the microphone and retell what you have heard from the lecture in your own words.

You will have 40 seconds to give your response.

## Assessment fairness

Not biased against any learner group

**Humans are great language assessors**  
but can also be (unconsciously) **biased**.

**AI is not distracted by irrelevant factors:**  
appearance, personality, body language.



## Assessment reliability

Consistent results across testing occasions

**Humans are great language assessors**  
but can also get **tired & inconsistent.**

**AI will rate the same work similarly,**  
regardless of time, and is **always available.**



# AI assessment requires rigorous AI model development & validation

**VERSANT™**  
John Smith

Versant Professional English Test - Level 2

Test Completion Date (GMT): 01 January 2021  
Test Identification Number (TIN): 12345678

**Overall GSE Score**  
**66**  
CEFR: B2

The candidate can use a good range of vocabulary, collocations and functions and can express ideas and opinions on some abstract topics. They can summarize, comment on and discuss a wide range of factual and imaginative texts. They can understand the speaker's point of view on most topics delivered at natural speed and in standard language. They can present clear, detailed information and points of view on a wide range of familiar topics. They can develop a clear description or narrative and supporting main points with relevant detail.

**Understanding the Skills**  
**Overall Score**  
The Overall score on this test reflects a candidate's ability to understand spoken and written English in the international workplace. To get a high score, candidates need to be able to respond appropriately in various spoken and written tasks. Speaking, listening, reading, writing, grammar, and vocabulary are assessed.

**Current Capabilities in Detail**

**Speaking: GSE: 63/90** CEFR: B2

The candidate can generally communicate with accuracy on most topics and in most contexts. They may make errors when talking about very unfamiliar topics but these do not prevent understanding. They can contribute to a group conversation fluently and naturally, provided the topic is not too complex. They can perform a range of work-related tasks such as discussing creative ideas for work-related projects, participating effectively in meetings, describing a product or service concisely, and giving clear instructions about solving job-related problems. They can give a clear, detailed spoken description of how to carry out a procedure.

**Tips to improve:**

- Learn and use grammar, which can help you form opinions about a future event (e.g. modal verbs).
- Practice talking about an idea if the exact language is not known.
- Practice summarizing ideas by paraphrasing (repeating something using different words).
- Work on building vocabulary for effective participation in a discussion (e.g. phrases to gain time while thinking what to say, introducing a new topic/information, and developing a clear argument with supporting points and relevant examples).
- Practice giving and asking about opinions on various topics without too much preparation.

**Business Partner - B2:**  
SB: Unit 2.1 Project 8 SB: Unit 2.4 Listening 2  
SB: Unit 8.2 Speaking 9 SB: Unit 3.3 video 2-5  
SB: Unit 5.4 Speaking Task 4 SB: Unit 5.4 Listening 2  
SB: Unit 6.4 Task 4 SB: BW5 Changing Expectations 2-7

**Understanding the Skills**  
**Speaking**  
The Speaking score reflects the ability to communicate in a range of everyday and workplace situations. The score is based on the ability to produce fluent, intelligible speech by using appropriate stress, rhythm, and intonation as well as accurate or appropriate grammar.

**Listening: GSE: 58/90** CEFR: B1+

The candidate can understand most TV programs on familiar topics, structured presentations within their own field, and classes and training courses on a range of topics (provided the content is simple). They can guess the meaning of unknown words from the context if the discussed topic is familiar. They can recognize examples and their relation to the idea they support.

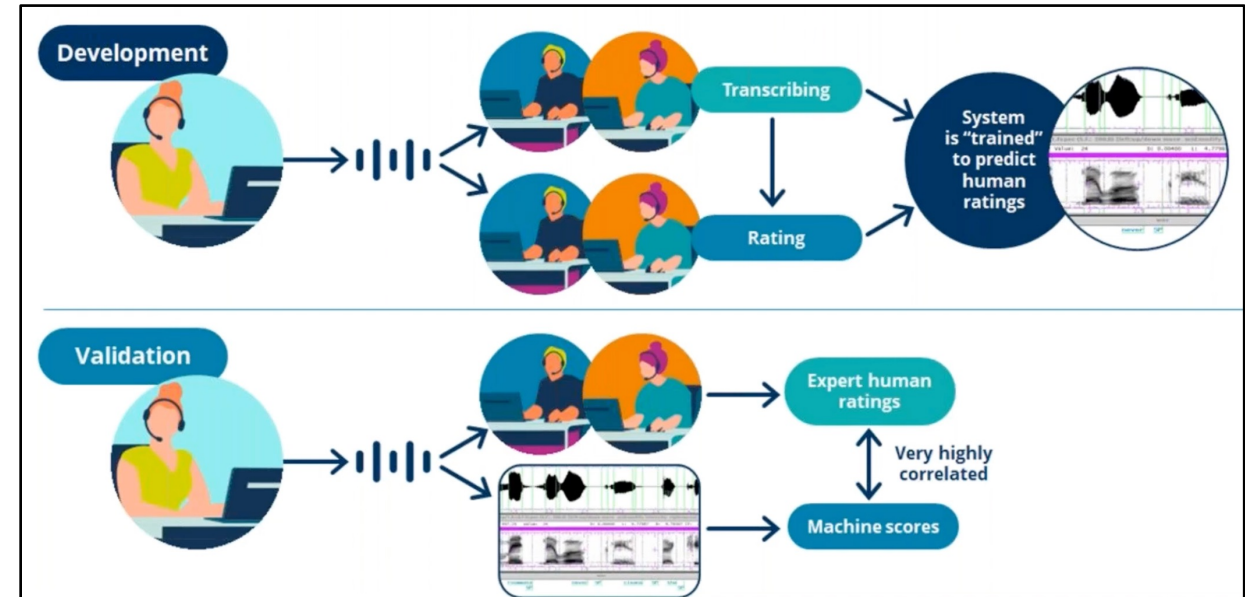
**Tips to improve:**

- Practice guessing the meaning of new words from the context.
- Listen to authentic materials at home (e.g. TV programs, presentations or discussions). Try to follow the speech (even if it is fast) and understand what the speaker(s) says.
- Practice identifying different strategies that speakers use in a conversation or discussion (e.g. rhetorical questions, a joke, agreeing or disagreeing, offering an example or solutions).
- Try to recognize the difference between a range of views and opinions when listening to discussions.

**Business Partner - B2:**  
SB: Unit 2.4 Listening 2 SB: Unit 3.3 video 2-5  
SB: Unit 3.3 video 2-5 SB: Unit 3.3 video 2-5  
SB: Unit 5.4 Listening 2 SB: BW3 Investment opportunities 2-3  
SB: BW5 Changing Expectations 2-7

**Understanding the Skills**  
**Reading**  
The Reading score reflects the ability to understand written English in the international workplace. The score is based on the ability to understand written English in the international workplace. The score is based on the ability to understand written English in the international workplace. The score is based on the ability to understand written English in the international workplace.

**Writing**  
The Writing score reflects the ability to produce written English in the international workplace. The score is based on the ability to produce written English in the international workplace. The score is based on the ability to produce written English in the international workplace. The score is based on the ability to produce written English in the international workplace.

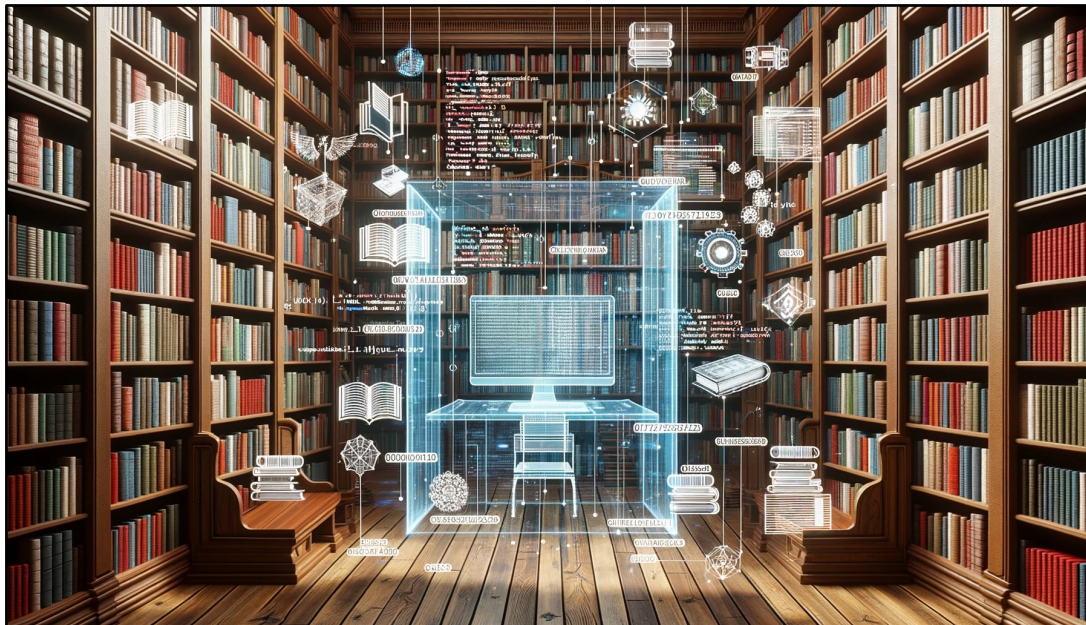




# Large Language Models learn *without* human supervision

AI reads **vast Internet datasets** to learn how to imitate human language & tasks.

Humans evaluate model's capabilities and **try to align it** to expectations.





# Out of the box, current-generation LLMs are unsuited for summative high-stakes assessment

- **Their output cannot be 100% reliable.** The same essay will be scored differently, even with clear rubrics.
- Their complexity makes it **challenging to finetune their behaviour**, esp. for proprietary & closed LLMs.
- They are trained on datasets that can **perpetuate socio-cultural biases**, esp. in English-first models.
- While they can generate feedback, extra attention needs to be paid to **proper safeguarding**.

## Coherence: 9/10

- The essay is logically organized, with clear subdivisions for each point and a well-written introduction and conclusion. The points flow logically from one to the next, providing a coherent and comprehensive argument. Some transitions between points could be smoother.

## Coherence: 8/10

The essay is logically structured, with clear and coherent arguments. Each point is explained in a separate paragraph, and the use of headings and bullet points makes it easy to follow. There is a slight repetition in the last two paragraphs, which could be avoided to make the essay more concise and effective.

However, LLMs present opportunities for formative assessment

1. Decrease **time & cost of building less rigorous** AI systems, e.g. test prep.
2. Facilitate **development of *future skills***, e.g. collaboration, cultural intelligence.
3. Enable new types of **immersive learning experiences**, e.g. virtual worlds.



The advertisement features a smartphone screen on the left showing a conversation with an AI character named LUNA. The screen displays the text: "Hello, I'm LUNA! It's nice to meet you. How are you today?" and "Excellent! How about you?". To the right of the screen, the text reads: "Build your confidence in speaking English. Be the first to meet Mondly by Pearson's new AI conversation partner." The Mondly by Pearson logo is in the top right corner.

Pearson's [Skills Outlook](#) report:



### Looking ahead

The 10 skills needing most improvement to meet the demands of the country's economy by 2026 are:

- |    |                                |    |                       |
|----|--------------------------------|----|-----------------------|
| 01 | Collaboration                  | 06 | Agility               |
| 02 | Customer Focus                 | 07 | Emtional Intelligence |
| 03 | Personal Learning & Mastery    | 08 | People Management     |
| 04 | Achievement Focus              | 09 | Communication         |
| 05 | Cultural & Social Intelligence | 10 | Direction & Purpose   |

# Future of generative AI depends on ongoing copyright lawsuits

2015: Google wins vs Authors Guild.  
Turning copyrighted books to create a searchable database is **fair use**.

This recognition of the law's history in comparative context should compel policymakers, who might be inclined to discount this public interest dimension in current copyright debates, to actively distinguish today's circumstances from that of the past. They, along with any interest groups or lobbyists that promote a particular copyright agenda, would have to clearly and cogently 'make their case' for derogating from this policy dimension or any other 'transcendent' principle, at least in those jurisdictions that have been the subject of close historical scrutiny. The more that scholars engage in uncovering the many copyright 'origin stories' around the world, the more a global picture will emerge that will provide greater depth of understanding on the comparative aspects of the law, including any areas of convergence or divergence.

Pages 46 to 128 are not shown in this preview.

2023: similar problem, similar defence, similar resolution?

*The Atlantic*

## REVEALED: THE AUTHORS WHOSE PIRATED BOOKS ARE POWERING GENERATIVE AI

Stephen King, Zadie Smith, and Michael Pollan are among thousands of writers whose copyrighted works are being used to train large language models.

By Alex Reisner

*Editor's note: This article is part of The Atlantic's series on Books3.*



**AG** THE Authors Guild

Send a Letter to AI Companies Telling Them They Do Not Have the Right to Use Your Work

TELL AI COMPANIES THEY DO NOT HAVE THE RIGHT TO USE YOUR WORK

[SEND A LETTER NOW](#)

# Some LLM vendors will now protect users against:

1. Copyright claims regarding model training data
2. Unintended copyright infringements in generated output

## Microsoft announces new Copilot Copyright Commitment for customers

Sep 7, 2023 | Brad Smith, Vice Chair and President, Hossein Nowbar, CVP and Chief Legal Officer

Microsoft's AI-powered Copilots are changing the way we work, making customers more efficient while unlocking new levels of creativity. While these transformative tools open doors to new possibilities, they are also raising new questions. Some customers are concerned about the risk of IP infringement claims if they use the output produced by generative AI. This is understandable, given recent public inquiries by authors and artists regarding how their own work is being used in conjunction with AI models and services.

## Shared fate: Protecting customers with generative AI indemnification

October 13, 2023

**Neal Suggs**  
VP Legal, Google Cloud

**Phil Venables**  
VP, TI Security & CISO, Google Cloud

### To our customers:

At Google Cloud, we put your interests first. This means that when you choose to work with us, we become partners on a journey of shared innovation, shared support, and shared fate.

# Model output copyright: uncharted territory?



## Images:

- US copyright law applies to humans, **not animals/machines**
- Text-to-image generators are not **tools** that can be **used** by author
- Sufficient **human authorship**: AI comic book *yes*, just images *no*
- Copyright Office **won lawsuits** but **court admitted new laws needed**



## Text & code:

- **As above**, verbatim LLM output likely goes into public domain
- Can be granted copyright **only** after proving significant edits



# EU AI Act proposes a *risk-based* safety framework

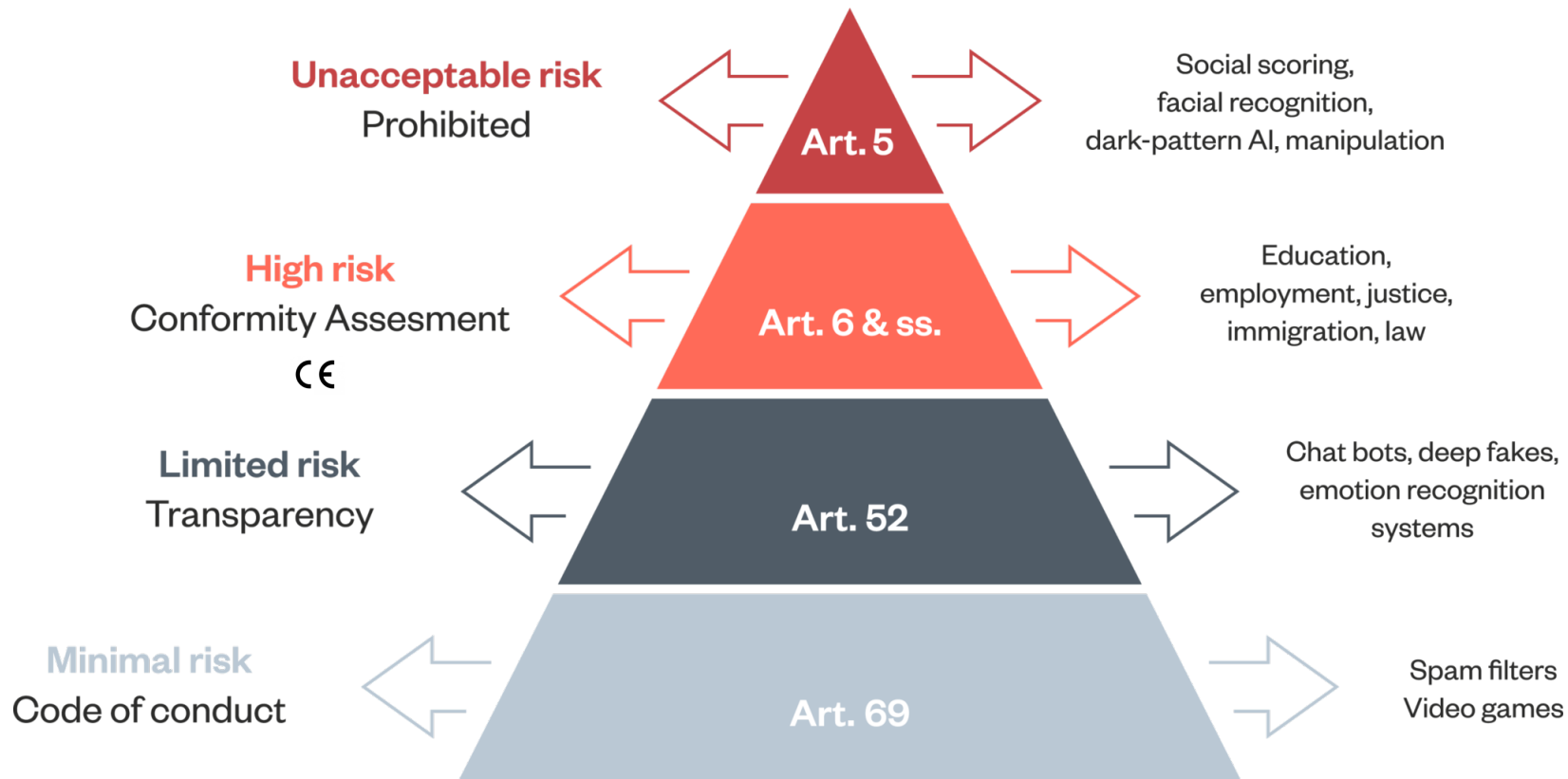


Image: Lilian Edwards (Ada Lovelace Institute)

Pre/post-market requirements for high-risk system providers:

- training data governance,
- technical documentation,
- risk management,
- human oversight,
- accidents reported to authorities.

Significant model finetuning can make you the provider.

Open-source models more likely to meet obligations?



# E-LEARNING FUSION 2023

Organizator:

Digital Learning Centre



Let's Fintech

